Dr. Brij Mohan, 2026, 14:4
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Text Summarization for News Articles Using an AI Chatbot

Dr. Brij Mohan[1], Sagar Chaudhary[2]
[1]Director of QST Dept. Quantum University, Roorkee, India
[2]Assistant Professor Quantum University, Roorkee, India

Abstract- Massive amounts of textual content every day are produced due to the rapid expansion of digital news channels. Quickly drawing significant conclusions from lengthy news articles may be a challenge for readers, journalists, and academics. The problem of condensing long articles into brief and instructive summaries has been solved by automating text summarizing, which preserves the fundamental meaning and context. The design and implementation of an AI chatbot-based system for automatic text summarization of news articles are presented in this research paper using deep learning and NLP approaches. The suggested system provides support for abstract, hybrid, and extractive summarization techniques. Despite the sophistication of transformer-based models like BERT[4, T5[7], and PEGASUS[7] Abstract summarizing is done by using traditional extractive techniques like TF- to provide summaries that are human-like. The statistical significance of IDF[2] and TextRank[8] is used to identify significant phrases. The modular system architecture includes preprocessing, summary engine, evaluation module, and interactive chatbot interface, making it easy for users to enter news items and receive summarized results instantly.

Keywords: Text Summarization, AI Chatbot, Natural Language Processing, News Articles, Transformer Models, Deep Learning.

I. INTRODUCTION

The internet's general availability and rapid development of digital technology have led to an exponential increase in the production and consumption of online news information. Blogs, social media sites, internet portals, and newspapers post thousands of news stories every hour. The abundance of information offers a guarantee of greater transparency and knowledge, but it also poses a significant problem known as information overload. It can be hard for readers to find reliable and relevant information in a short amount of time when news stories are long and cover various viewpoints, facts, and opinions. As a result, there is a growing demand for intelligent systems that can automatically reduce news stories into brief, insightful summaries without sacrificing important information

Text summarizing is a significant use of NLP, which aims to produce a condensed version of a longer document while maintaining its main points, context, and factual accuracy. Summarization is vital in the news industry for academics to comprehend vast amounts of textual data, journalists to efficiently analyze different sources, and readers to quickly grasp significant events. When there is a need for immediate access to information, such as breaking news, financial updates, and political reporting, Automated summary systems are very helpful.

Traditional text summarization techniques primarily used statistical features like word frequency, sentence position, and similarity measures to select important sentences from the original text. Despite their computational efficiency and preservation of factual accuracy, these methods often produce summaries that lack coherence and readability. Summarization systems have undergone significant evolution with the advent of machine learning and deep learning approaches, specifically sequence-to-sequence models and attention mechanisms.

Text summarization has been transformed by recent developments in transformer-based architectures, such as BERT[4], T5[7], BART[5], and BART[5].

Models can comprehend long-range connections and contextual linkages within text through PEGASUS[6]. Abstract summarization is made possible by these models by creating new sentences

that accurately convey the article's main ideas in a way that is similar to that of a person. Abstract models can be helpful, but they also have drawbacks like bias, hallucinations, and high processing costs.

Summarization systems can be deployed in this area using AI chatbots, which have become an interactive and user-friendly interface. News stories can be inputted by users.

By combining summarizing models and conversational agents, they can adjust the length of their summaries and obtain instant results. This project aims to develop an AI chatbot-based text summary system for news stories that balances accuracy and fluency by combining extractive and abstractive methods. The suggested solution aims to reduce information overload, improve news accessibility, and demonstrate the practical application of cutting-edge NLP approaches in everyday situations.

II. RELATED WORK

Statistical and rule-based techniques, such as sentence position and word frequency, were the focus of early text summarization studies. Graph-based algorithms for extractive summarization, such as TextRank[8] and LexRank, gained prominence later. High-quality abstractive summarization is now possible with the advent of deep learning through the use of sequence-to-sequence models and transformer architectures.

III. PROPOSED METHODOLOGY

The suggested approach for utilizing an AI chatbot to summarize news stories is designed to be modular, scalable, and flexible enough to meet different summarization needs. The system combines machine learning and deep learning models with NLP approaches to create summaries that are succinct, logical, and context-aware. Data collection, preprocessing, summarization, post-processing, evaluation, and chatbot are all tasks that involve data collection, preprocessing, summarization, post-processing, evaluation, and chatbot. Interactions are a part of the technique's sequential steps.

The first phase involves data collection and input management. The system accepts news articles in multiple formats, including document files like PDF or TXT, URLs of online news portals, and raw text input. The primary article content is collected from URL-based inputs using web scraping techniques, which eliminates unnecessary metadata, navigation menus, and advertisements. The summarization models are also trained and validated using publicly accessible datasets such as CNN/Daily Mail and BBC News.

The preprocessing phase involves cleaning and normalizing the raw text to enhance model performance. The removal of HTML tags, unnecessary whitespace, and special characters leads to sentence segmentation and tokenization. When necessary, stop-words are reduced and normalized using strategies such as lowercaseing. To maintain essential factual information while summarizing, Named Entity Recognition (NER) can be employed to recognize significant entities, such as names, locations, dates, and organizations.

Hybrid, abstractive, and extractive summarization are all supported by the core summarization stage. TF-IDF and TextRank[8] are used as statistical and graph-based algorithms.

Extractive summarization is used to rank and score phrases based on their significance and relevance. The top-ranked sentences are selected with the least amount of repetition to give a succinct summary. By selecting specific sentences from the original article, our method ensures factual accuracy.

Deep learning models that are based on transformers, including BERT[4], T5[7], BART[5], and Abstract summarization is done using PEGASUS[6]. The whole text is encoded by these models and new sentences are produced that convey the overall meaning of the news article in a human-like manner. To increase coherence and fluency, strategies such as length penalties, beam search, and attention processes are used. A hierarchical approach is used

to break down long news stories into digestible sections and summarize them.

The hybrid summarizing approach combines the benefits of both extractive and abstractive approaches. Extractive approaches are used to find significant sentences that are then paraphrased and enhanced using an abstract model. Reading and contextual flow are preserved while hallucinations are reduced.

The generated summaries are evaluated using common automatic metrics such as ROUGE[3], BLEU[10], and BERTScore during the review phase. By comparing, these metrics evaluate content overlap, semantic similarity, and fluency System-generated summaries with reference summaries. Humans evaluate relevance, coherence, grammatical accuracy, and factual consistency.

The AI chatbot interface represents the user interaction layer finally. Through a conversational interface, users may submit news stories, select the type and length of summary, and receive results instantly. This integration improves usability and illustrates how the suggested methodology can be applied practically in actual news consumption situations.

IV. SYSTEM ARCHITECTURE

The system architecture of the proposed AI chatbot-based text summarization framework uses a modular and layered approach to ensure scalability, flexibility, and effective news item processing. Assessment methods, NLP-based summarization models, backend processing, and user interaction are all integrated into a single pipeline by the system.

An AI chatbot interface at the front-end layer is used by users to interact with the system. This interface allows users to select the type and length of a summary and submit news stories in the form of raw text, URLs, or uploaded files. The system is accessible to non-technical individuals.

The chatbot ensures a conversational and easy-to-use interface.

The backend layer is the central controller of the system. Through APIs that are clearly defined, it manages user requests, validates inputs, and coordinates communication between various processing modules. The purpose of this layer is to ensure that summarizing duties are handled safely, reliably, and on time.

The preprocessing module cleans and standardizes the input news content. Performing tokenization, stop-word removal, sentence segmentation, and optional named entity recognition is necessary to preserve crucial factual information. The performance of both extractive and abstractive summarization models is significantly improved after preprocessing.

The summarization layer is comprised of abstractive and extractive engines. The abstractive summarizer employs transformer-based models to give fluent and context-aware summaries, while the extractive summarizer utilizes statistical and graph-based techniques to recognize significant sentences. A hybrid strategy is employed to enhance readability and factual accuracy by combining both techniques. Finally, ROUGE[3], BLEU[10], and The assessment module evaluates the produced summaries using BERTScore metrics. The chatbot's user interface receives the final summary after that.

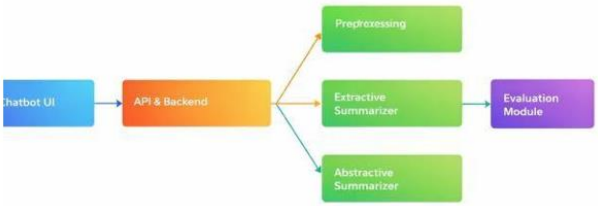

Figure 1: The system architecture of the proposed AI chatbot-based text summarization system.

V. RESULTS AND DISCUSSION

This section addresses the effectiveness of various summarizing techniques and provides a thorough analysis of the experimental findings from the suggested AI chatbot-based text summarization system. The method was assessed utilizing a combination of qualitative human evaluation and

automatic evaluation measures to guarantee the generated summaries' practical usability as well as their numerical dependability.

Standard benchmark news datasets like CNN/DailyMail and BBC News, which include a variety of articles from a variety of fields like politics, business, sports, and technology, were used for the experimental evaluation. These datasets offer trustworthy reference summaries for comparison and are frequently utilized in summarizing studies. To achieve a fair comparison, the system was tested under similar preprocessing and input settings using extractive, abstractive, and hybrid summarizing strategies.

Techniques like TextRank[8] and TF- IDF[2] were assessed for extractive summarization. Due to their direct selection of lines from the source news stories, these methods showed excellent effectiveness in maintaining factual information. Good content overlap with reference summaries was shown by the ROUGE-1 and ROUGE-2 ratings for extractive approaches, which were consistently constant. However, because the chosen sentences were not always logically connected, qualitative examination showed that extractive summaries frequently lacked coherence and smooth flow. Overall readability was impacted by this restriction, particularly for lengthier pieces.

In terms of fluency and contextual understanding, abstractive summarization models based on transformer topologies, such BERT[4], T5[7], BART[5], and PEGASUS[6], demonstrated noticeably better performance. When compared to extractive techniques, these models' higher BERTScore and ROUGE-L scores demonstrate their capacity to capture semantic meaning as opposed to merely lexical overlap. Abstractive summaries received higher ratings for coherence, grammatical accuracy, and natural language quality from human evaluators. However, when dealing with lengthy or complicated news stories, abstractive models occasionally created little factual discrepancies or hallucinations.

The greatest balanced performance was obtained with the hybrid summarizing method. The hybrid approach effectively decreased duplication while preserving factual correctness by first identifying key sentences using extractive techniques and then honing them using abstractive models. This strategy improved readability and coherence while achieving competitive ROUGE[3] scores. The findings of human evaluation demonstrated that hybrid summaries were more fluent than extractive summaries and more dependable than exclusively abstractive summaries.

Response time and computing efficiency were also taken into account in performance analysis. Extractive methods are appropriate for low- resource contexts because they need little computer power and generate summaries fairly instantaneously. Conversely, transformer-based abstractive models produced better summaries but required more processing time and GPU support. By letting users choose the summary style according to their requirements, the AI chatbot interface successfully handled these trade-offs.

Overall, the findings show that transformer-based and hybrid summarization techniques are very successful at summarizing news articles in practical settings. By enabling interactive, customisable, and real-time summarization, the integration of these models with an AI chatbot greatly improves user experience. The debate emphasizes that although no single strategy is always the best, the suggested method offers adaptability and resilience, making it appropriate for a variety of news consuming situations.

V. CONCLUSION AND FUTURE WORK

This project successfully demonstrated the design and implementation of an AI chatbot-based text summarization system for news items in order to address the growing issue of information overload in the digital news ecosystem. The proposed approach combines machine learning and transformer-based deep learning models with Natural Language Processing (NLP) approaches to generate concise, intelligent, and context-aware summaries from

lengthy news items. The study evaluated the effectiveness of extractive, abstractive, and hybrid summarizing methods using both automated metrics and human evaluation.

The results of the experiment demonstrated that although extraction techniques such as TextRank and TF- IDF[2] are reliable and computationally effective in preserving factual content, they often lack coherence and natural language flow. Conversely, more readable, fluid, and semantically rich summaries were produced by transformer-based abstractive models like BERT[4], T5[7], BART[5], and PEGASUS [6], despite their higher processing cost and occasional factual inconsistencies. The hybrid technique turned out to be the most balanced solution, combining the factual dependability of extractive approaches with the language quality of abstractive models. The integration of these techniques into an AI chatbot interface significantly enhanced usability by allowing interactive, real-time, and customizable summaries.

All things considered, the proposed approach holds a lot of potential for real- world application in journalism, education, research, and digital media platforms. Future work may focus on integrating fact-checking and bias- detection techniques, enabling multilingual news summaries, and offering real-time summary using live news APIs in order to improve trustworthiness. Further improvements could include optimizing model efficiency for low-resource scenarios and improving the explainability of generated summaries. These improvements would contribute to the development of more reliable, scalable, and ethically sound AI-driven news summarization systems.

REFERENCES

1. I. Mani, Automatic Summarization, John Benjamins Publishing, 2001.
2. A. Nenkova and K. McKeown, "A survey of text summarization techniques," in Mining Text Data, Springer, 2012.
3. C.-Y. Lin, "ROUGE: A package for automatic evaluation of summaries," ACL Workshop, 2004.
4. J. Devlin et al., "BERT: Pre-training of deep bidirectional transformers for language understanding," NAACL-HLT, 2019.
5. M. Lewis et al., "BART: Denoising sequence-to-sequence pre-training for natural language generation," ACL, 2020.
6. J. Zhang et al., "PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization," ICML, 2020.
7. C. Raffel et al., "Exploring the limits of transfer learning with a unified text-to- text transformer," JMLR, 2020.
8. R. Mihalcea and P. Tarau, "TextRank: Bringing order into texts," EMNLP, 2004.
9. E. Barzilay and M. Elhadad, "Using lexical chains for text summarization," ACL Workshop, 1997.
10. K. Papineni et al., "BLEU: A method for automatic evaluation of machine translation," ACL, 2002.
11. H. P. Edmundson, "New methods in automatic extracting," Journal of the ACM, 1969.
12. R. Barzilay and M. Elhadad, "Using lexical chains for text summarization," ACL, 1997.
13. G. Erkan and D. R. Radev, "LexRank: Graph-based lexical centrality," JAIR, 2004.
14. K. Knight and D. Marcu, "Summarization beyond sentence extraction," ACL, 2000.
15. A. Rush et al., "A neural attention model for abstractive summarization," EMNLP, 2015.
16. R. Paulus et al., "A deep reinforced model for abstractive summarization," ICLR, 2018.
17. Y. Liu and M. Lapata, "Text summarization with pretrained encoders," EMNLP, 2019.
18. A. See et al., "Get to the point: Pointer-generator networks," ACL, 2017.
19. J. Cheng and M. Lapata, "Neural summarization by extracting sentences," EMNLP, 2016.
20. Z. Yang et al., "Hierarchical attention networks for document classification," NAACL, 2016.